

The clustering of galaxies at $z \sim 0.5$ in the SDSS-III Data Release 9 BOSS-CMASS sample: a test for the Λ CDM cosmology

S. E. Nuza^{1*}, A. G. Sánchez², F. Prada^{3,4,5,6}, A. Klypin⁷, D. J. Schlegel⁸, S. Gottlöber¹, A. D. Montero-Dorta⁵, M. Manera⁹, C. K. McBride¹⁰, A. J. Ross⁹, R. Angulo¹¹, M. Blanton¹², A. Bolton¹³, G. Favole⁵, L. Samushia⁹, F. Montesano², W. Percival⁹, N. Padmanabhan¹⁴, M. Steinmetz¹, J. Tinker¹², R. Skibba¹⁵, D. Schneider^{16,17}, H. Guo¹⁸, I. Zehavi¹⁸, Z. Zheng¹³, D. Bizyaev¹⁹, O. Malanushenko¹⁹, V. Malanushenko¹⁹, A. E. Oravetz¹⁹, D. J. Oravetz¹⁹, A. C. Shelden^{19,20}

¹ Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany

² Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstraße, 85748 Garching, Germany

³ Campus of International Excellence UAM+CSIC, Cantoblanco, E-28049 Madrid, Spain

⁴ Instituto de Física Teórica, (UAM/CSIC), Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain

⁵ Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain

⁶ Visiting scientist at the Max Planck Institut für Astrophysik (MPA), Karl-Schwarzschild-Str. 1, 85741 Garching, Germany

⁷ Astronomy Department, New Mexico State University, Las Cruces, NM, USA

⁸ Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, USA

⁹ Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth, PO1 3FX, UK

¹⁰ Department of Physics, Vanderbilt University, Nashville, TN, USA

¹¹ Max Planck Institut für Astrophysik (MPA), Karl-Schwarzschild-Str. 1, 85741 Garching, Germany

¹² Center for Cosmology and Particle Physics, New York University, NY, USA

¹³ Department of Physics and Astronomy, University of Utah, 115 south 1400 East, Salt Lake City, UT 84112, USA

¹⁴ Yale Center for Astronomy and Astrophysics, Yale University, New Haven, CT 06511, USA

¹⁵ Steward Observatory, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721, USA

¹⁶ Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA

¹⁷ Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA

¹⁸ Department of Astronomy, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA

¹⁹ Apache Point Observatory, P.O. Box 59, Sunspot, NM 88349-0059, USA

²⁰ Department of Astronomy, University of Florida, 211 Bryant Space Science Center, Gainesville, FL 326711-2055, USA

29 February 2012

ABSTRACT

We present results on the clustering of 282,068 galaxies in the Baryon Oscillation Spectroscopic Survey (BOSS) sample of massive galaxies with redshifts $0.4 < z < 0.7$ which is part of the Sloan Digital Sky Survey III project. Our results cover a large range of scales from $\sim 500 h^{-1} \text{ kpc}$ to $\sim 90 h^{-1} \text{ Mpc}$. We compare these estimates with the expectations of the flat Λ CDM standard cosmological model with parameters compatible with WMAP7 data. We use the MultiDark cosmological simulation, one of the largest N -body runs presently available, together with a simple halo abundance matching technique, to predict the galaxy correlation functions, power spectra, abundance of satellites and galaxy biases. We find that the Λ CDM model gives a reasonable description to the observed correlation functions at $z \sim 0.5$, which is a remarkably good agreement considering that the model, once matched to the observed abundance of BOSS galaxies, does not have any free parameters. However, we find a small ($\sim 10\%$) deviation in the correlation functions for scales around $10\text{--}30 h^{-1} \text{ Mpc}$. A more realistic abundance matching model and better statistics from upcoming observations

* E-mail: snuza@aip.de, arielsan@mpe.mpg.de, fprada@iaa.es

are needed to clarify the situation. We also predict that about 7% of the galaxies in the sample are most probably satellites inhabiting central haloes with mass $M \gtrsim 10^{14} h^{-1} M_{\odot}$. Using the MultiDark simulation we also study the scale-dependent galaxy bias b and find that $b \simeq 2$ for BOSS galaxies at scales $\gtrsim 10 h^{-1}$ Mpc. The large-scale bias, defined using the extrapolated linear matter power spectrum, depends on the maximum circular velocity V_{\max} of galaxies as $b = 1 + (V_{\max}/361 \text{ km s}^{-1})^{4/3}$, or on the galaxy number density n_g as $b = 0.0377 - 0.57 \log_{10} (n_g/h^3 \text{ Mpc}^{-3})$. The damping of the BAO signal produced by non-linear evolution leads to ~ 2 –4% dips in the large-scale bias factor defined in this way. Very accurate fits as a function of abundance and maximum circular velocity of galaxies are provided.

Key words: cosmology: large-scale structure of the Universe – cosmology: theory – galaxies: general – methods: observational – methods: numerical

1 INTRODUCTION

The clustering of galaxies is a fundamental measure of the statistical properties of the cosmic density field through cosmic time. In the last decade, it became possible to determine the clustering strength of galaxy populations at spatial scales out to tens of Mpc and beyond with reasonable accuracy by means of massive galaxy surveys such as the Two-Degree Field Galaxy Redshift Survey (e.g., Colless et al. 2001) and Sloan Digital Sky Survey (SDSS-I/II; e.g., Gunn et al. 1998; York et al. 2000; Gunn et al. 2006). These and previous studies have shown that the correlation function is not a simple power-law and that the correlation length of luminous and massive galaxies is larger than that of less luminous ones (see Zehavi et al. 2011, and references therein). Furthermore, it has been also shown that the clustering strength of Luminous Red Galaxies (LRGs) is an excellent tracer of the Baryon Acoustic Oscillation (BAO) signal, which can be used to constrain the expansion history of the Universe (e.g., Eisenstein et al. 2005).

The Baryon Oscillation Spectroscopic Survey (BOSS), a branch of the ongoing SDSS-III (Eisenstein et al. 2011), is considerably increasing the size of available galaxy samples. BOSS consists of galaxy and quasar spectroscopic surveys over a sky area of $10,000 \text{ deg}^2$ and its main goal is to measure the BAO feature at high precision. Specifically, BOSS aims at measuring the redshifts of about 1.5 million galaxies out to $z = 0.7$. It will also acquire about 150,000 Ly α forest spectra of quasars in the range $2.2 < z < 4$, to map the large-scale distribution of galaxies at these earlier epochs (see Slosar et al. 2011). The effective volume of the galaxy survey is expected to be about 7 times higher than that of the SDSS-I/II LRG sample which consisted of $\sim 100,000$ galaxies out to $z = 0.45$. The selection criteria of the BOSS targets results in a sample of massive, and hence highly clustered systems, which are suitable candidates for a reliable detection of the acoustic peak. Additionally, the project also provides a wealth of other information on clustering and physical properties of galaxies.

Requirements for theoretical predictions of galaxy clustering in BOSS are extreme: one needs accurate predictions for very large volumes in order to compare with observations. Therefore, the combination of large-volume cosmological N -body simulations with prescriptions to associate galaxies with dark matter haloes turns out to be the most efficient way to generate the required model galaxy sam-

ples. Recently, White et al. (2011) presented clustering results for scales in the range ~ 0.5 – $20 h^{-1}$ Mpc based on $\sim 44,000$ galaxies in the redshift range $0.4 < z < 0.7$ obtained during the first semester of BOSS operation. To compare these observational results with theory, the authors combined large, albeit low-resolution, N -body simulations with the Halo Occupation Distribution (HOD) model (e.g., Berlind & Weinberg 2002; Kravtsov et al. 2004; Zentner et al. 2005; Skibba & Sheth 2009; Ross & Brunner 2009; Ross et al. 2010). Their results suggest that the majority of BOSS galaxies are central systems living in haloes with a mass of $\sim 10^{13} h^{-1} M_{\odot}$, while about 10% of them are satellites typically residing in haloes ~ 10 times more massive.

The HOD approach is the most often used framework to make predictions for the large-scale distribution of galaxies. Alternatively, HODs can also be measured in observations (Zehavi et al. 2005; Abazajian 2005; Brown et al. 2008; Zheng et al. 2009). The main component of classical HOD models is the probability, $P(N|M)$, that a halo of virial mass M hosts N galaxies with some specified properties. In general, theoretical HODs require the fitting of a function with several parameters (e.g., Kravtsov et al. 2004; Zheng et al. 2005), which gives some freedom to match the observed clustering of galaxies. These models also depend on the theoretical approach adopted to predict the galaxy number N inside haloes of mass M . For example, Zheng et al. (2005) used SPH simulations and semi-analytical models to measure the number of galaxies as a function of hosting halo mass, which is definitely a challenging theoretical exercise. White et al. (2011) tuned five HOD free parameters to fit the observed clustering of galaxies. In this case a random fraction of dark matter particles is selected from the simulations with a fraction following the optimized HOD. This prescription will have the best match to observations hence producing good-quality mock catalogs. However, this is not the best way of testing a cosmological model. Kravtsov et al. (2004) used a different approach: they identify subhaloes in high-resolution N -body simulations in order to associate them with satellite galaxies. This is a more attractive path, which can be further perfected by more accurate simulations and more elaborate prescriptions for “galaxies” in dark matter-only simulations (e.g., Trujillo-Gomez et al. 2011).

Halo Abundance Matching (HAM) has recently emerged as an attractive alternative to HOD in or-

der to bridge the gap between dark matter haloes and galaxies (Kravtsov et al. 2004; Tasitsiomi et al. 2004; Vale & Ostriker 2004; Conroy et al. 2006; Guo et al. 2010; Wetzel & White 2010; Trujillo-Gomez et al. 2011). Abundance-matching resolves the issue of connecting observed galaxies to simulated dark matter haloes and subhaloes by setting a one-to-one correspondence between the red-band luminosity or stellar and dynamical masses: more luminous galaxies are assigned to more massive (sub)haloes. By construction, the method reproduces the observed luminosity function (or stellar mass function). It also reproduces the scale dependence of galaxy clustering over a large range of epochs (Conroy et al. 2006; Guo et al. 2010). When abundance matching is used for the observed stellar mass function (Li & White 2009), it gives also a reasonable fit to lensing results (Mandelbaum et al. 2006) and to the relation between stellar and virial mass (Guo et al. 2010). Guo et al. (2010) also attempted to reproduce the observed relation between the stellar mass and the maximum circular velocity with partial success, finding deviations both in shape and amplitude between predictions and observations. At circular velocities in the range $100\text{--}150\text{ km s}^{-1}$ the predicted circular velocity was $\sim 25\%$ lower than the observed one. They proposed that this disagreement is likely due to the fact that they did not include the effect of baryons. Indeed, Trujillo-Gomez et al. (2011) show that accounting for baryons drastically improves the situation.

Just like as with HODs, there are different flavours of HAMs. Generally, one does not expect a pure monotonic relation between stellar and dynamical masses. There should be some degree of stochasticity in this relation due to deviations in the merger history, angular momentum, and halo concentration. Even for haloes (or subhaloes) with the same mass, these properties should be different for different systems, which would lead to deviations in stellar mass. Observational errors are also responsible in part for the non-monotonic relation between halo and stellar masses. Most of modern HAM models already implement prescriptions to account for the stochasticity (Tasitsiomi et al. 2004; Behroozi et al. 2010; Trujillo-Gomez et al. 2011; Leauthaud et al. 2011). The difference between monotonic and stochastic models depends on the magnitude of the scatter and on the stellar mass. The typical value of the scatter in the r -band is expected to be $\Delta M_r = 0.3\text{--}0.5\text{ mag}$ (e.g., Trujillo-Gomez et al. 2011). For the Milky-Way-size galaxies the differences are practically negligible (Behroozi et al. 2010), but they increase for very massive galaxies such as those targeted with BOSS due to the strong dependence of the bias with mass.

Almost two years after the start of the project, BOSS has obtained the spectra of about 487,000 galaxies and 61,000 quasars. Using the SDSS-III Data Release 9 (DR9) BOSS data we present results on the two-dimensional, projected and redshift-space correlation functions on scales from $\sim 500 h^{-1}\text{ kpc}$ to $\sim 90 h^{-1}\text{ Mpc}$ including fibre collision corrections. In order to make predictions for the ΛCDM cosmological model we use a large high-resolution N -body simulation with a resolution high enough to resolve subhaloes, which is very important for the HAM prescription. When connecting haloes with galaxies we use a stochastic HAM model.

This paper is organized as follows. In Section 2 we

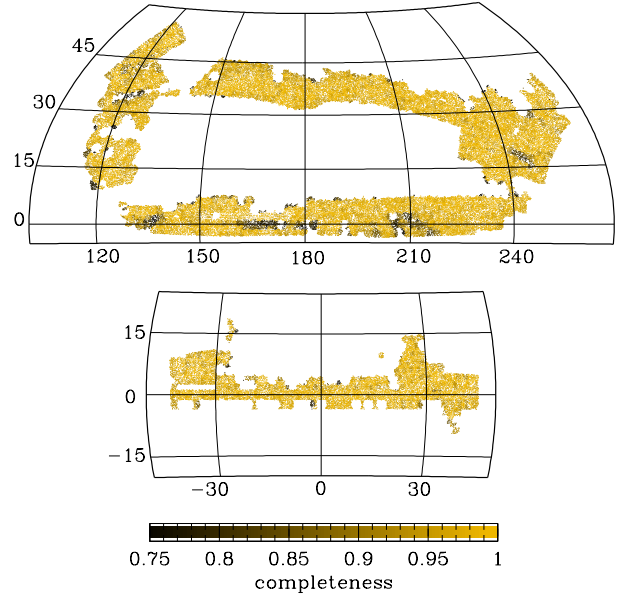


Figure 1. Sky area covered by the DR9 BOSS-CMASS sample shown in Aitoff projection colour-coded by completeness (see text). The upper and lower maps display the northern and southern galactic caps respectively.

present the BOSS galaxy sample studied here, dubbed “CMASS”, and the measurements of the two-dimensional, projected and redshift-space galaxy clustering in observations. In Section 3 we present the details of the Multi-Dark simulation, the halo catalogs and the HAM technique adopted here. In Sections 4 and 5 we compare the clustering measures with observations and study the occupation distribution given by our halo catalog. We also discuss the comparison between our halo occupation distribution with that obtained by White et al. (2011) using an HOD model. In Section 6 we study the scale-dependent bias of galaxy clustering of the CMASS sample as inferred from our HAM model both in real and Fourier space. Finally, in Section 7 we close the paper with the summary and conclusions.

In Appendix A we discuss several effects that can affect the clustering power.

2 OBSERVATIONS

2.1 The CMASS sample

In this section we introduce the BOSS sample of massive galaxies analyzed in this work. The target galaxies are selected in such a way that the stellar mass of the systems is approximately constant over the entire redshift range of interest. As a consequence, the resulting galaxy sample is usually dubbed ‘constant mass’ (CMASS) sample. These galaxies are characterized by high-luminosities which translate in a rather low comoving space density of about $3 \times 10^{-4} h^{-1}\text{ Mpc}$. The sample can be obtained by applying the following colour cuts to the observations (see e.g. Eisenstein et al. 2011):

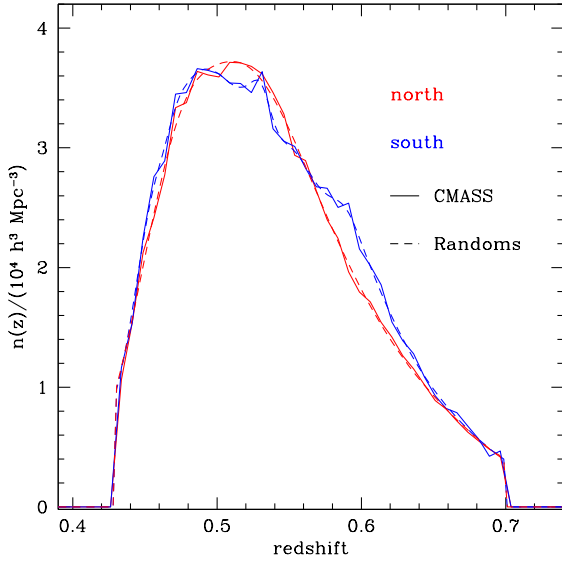


Figure 2. The comoving number density of galaxies in the DR9 BOSS-CMASS sample both for the north and south subsamples in the redshift range $0.4 < z < 0.7$. Dashed lines show the smoothed distributions used to create the Poisson distribution of particles when computing the correlation functions (see text).

$$\begin{aligned}
 17.5 &< i_{\text{cmod}} < 19.9, \\
 r_{\text{mod}} - i_{\text{mod}} &< 2, \\
 i_{\text{fiber2}} &< 21.5, \\
 d_{\perp} &> 0.55, \\
 i_{\text{cmod}} &< 19.86 + 1.6 \times (d_{\perp} - 0.8)
 \end{aligned} \tag{1}$$

where $d_{\perp} = (r_{\text{mod}} - i_{\text{mod}}) - (g_{\text{mod}} - r_{\text{mod}})/8$ and i_{fiber2} is the i magnitude measured with the $2''$ BOSS fiber within the SDSS *ugriz* photometric system (Fukugita et al. 1996). The subscripts cmod and mod denote “cmodel” and “model” magnitudes respectively. These cuts are chosen to pick out massive red galaxies at $z \gtrsim 0.4$. In particular, the condition $d_{\perp} > 0.55$ selects systems with observed red $r - i$ colours, whereas the conditions imposed on the i -magnitude is designed to identify an approximately complete galaxy sample down to a limiting stellar mass. Most of these galaxies ($\sim 75\%$) show an early-type morphology with a characteristic stellar mass of $M_{*} \sim 10^{11} h^{-1} M_{\odot}$ and an absolute r -band magnitude of $M_r - 5 \log h \lesssim -20.7$ (Masters et al. 2011).

Schlaflly et al. (2010) and Schlaflly & Finkbeiner (2011) found systematic offsets between the colours of SDSS objects in the southern and northern Galactic hemispheres which might reflect a combination of percent calibration errors in the SDSS photometry and errors in the corrections for Galactic extinction. The Schlaflly & Finkbeiner (2011) results suggest a systematic offset in the value of d_{\perp} of 0.0064 between the north and south. As the CMASS selection criteria depends on d_{\perp} , this offset leads, in principle, to a difference in the galaxy samples selected for spectroscopic observations in the two hemispheres. Ross et al. (2011) found a 2% difference in the number density of CMASS targets

between the northern and southern hemispheres, which reduces to 0.3% when this offset is applied to the galaxies in the south before applying the CMASS selection criteria. However, Ho et al. (2012) found no appreciable north/south colour offsets in their sample. In this work we do not apply a colour offset to the selection of CMASS galaxies in the south. Although we present results obtained from the combined (north+south) CMASS sample, we also analyse the data from the northern and southern hemispheres separately in order to avoid potential systematics that could be associated with the use of slightly different selection criteria.

For a number of reasons it is not possible to obtain reliable redshifts for all the galaxies satisfying the CMASS selection criteria (see Section 2.2). We estimate the completeness $c = n_z/n_t$, where n_t is the number of galaxy targets and n_z the number of these with reliable redshift estimates (weighted as described in Section 2.2) for each sector of the survey mask, that is, the areas of the sky covered by a unique set of spectroscopic tiles (Blanton et al. 2003; Tegmark et al. 2004) which we characterize using the MANGLE software (Hamilton & Tegmark 2004; Swanson et al. 2008). The average completeness of the combined CMASS sample is 98.2%. We trim the final area of our sample to all sectors with completeness $c \geq 0.75$, producing our final sample of 282,068 galaxies, of which 219,773 and 62,295 are located in the northern and southern galactic caps respectively. Fig. 1 shows an Aitoff projection of the resulting survey mask in the northern (upper panel) and southern (lower panel) regions, with effective areas $A_{\text{eff}} = \sum_i c_i \Omega_i$, where the sum extends over all sectors contained in the mask and Ω_i corresponds to their solid angles, 2502 deg² and 688 deg² respectively. The redshift distribution of the CMASS sample can be seen in Fig. 2 both for the northern and southern subsamples. The dashed lines show the smoothed distributions used to create the random samples of points for our clustering analysis (see Section 2.2). As shown in this figure the galaxy number density peaks at $z \simeq 0.52$ having a value of $\bar{n}_g \simeq 3.6 \times 10^{-4} h^3 \text{ Mpc}^{-3}$ and a mean redshift of $z = 0.55$.

2.2 Clustering measures

We characterize the clustering of the CMASS galaxy sample by means of two-point statistics in configuration space. We measure the angle-averaged redshift-space correlation function $\xi(s)$ and the full two-dimensional $\xi(\sigma, \pi)$, where σ and π are the components in the direction perpendicular and parallel to the line of sight of the total separation vector \mathbf{s} . These measurements are affected by redshift-space distortions. In order to obtain a clustering measure that is less sensitive to these effects we also compute the projected correlation function (Davis & Peebles 1983)

$$\Xi(\sigma) = 2 \int_0^{\infty} \xi(\sigma, \pi) d\pi. \tag{2}$$

In practice, we sum all pairs with $\pi_{\text{max}} < 100 h^{-1} \text{ Mpc}$.

We compute the full correlation functions $\xi(\sigma, \pi)$ using the Landy & Szalay (1993) estimator

$$\xi(\sigma, \pi) = \frac{DD - 2DR + RR}{RR} \tag{3}$$

where DD , DR , and RR are the suitably normalized numbers of data-data, data-random, and random-random pair

counts in each bin of (σ, π) . In order to measure these quantities without introducing systematic effects, a few important corrections must be taken into account. Here we give a brief description of the main issues that should be considered while a more detailed discussion will be presented in Ross et al. (2012).

As described in the previous section, the spectroscopic CMASS sample is constructed from a target list drawn from the SDSS photometric observations. Even though the overall completeness of the CMASS sample is high, it is not possible to obtain reliable redshifts for all galaxies satisfying the selection criteria specified in Section 2.1. Which galaxies are observed spectroscopically is determined by an adaptive tiling algorithm, based on that of Blanton et al. (2003), which attempts to maximize the number of measured spectra over the survey area. As a result of this algorithm, not all galaxies satisfying the CMASS criteria are selected as targets for spectroscopy. Even when a fibre is assigned to a galaxy and a spectrum is observed, it might not be possible to obtain a reliable estimation of the redshift of the object, leading to what is called a *redshift failure*. These tend to occur for fibres located near the edges of the observed plates. This implies that it is not possible to simply consider these redshift failures as an extra component affecting the overall completeness of the sector since their probability is not uniform across the field. In order to correct for this effect we define a set of weights, w_{zf} , whose default value is one for all galaxies in the sample. For every galaxy with a redshift failure, we increase by one the value of w_{zf} of the nearest galaxy with a good redshift measurement. The application of these weights effectively corrects for the non-uniformity effects produced by redshift failures.

The main cause for the loss of objects is, however, fibre collisions (Zehavi et al. 2002; Masjedi et al. 2006). The BOSS spectrographs are fed by optical fibres plugged on plates, which must be separated by at least $62''$ (in the concordance cosmology this corresponds to a distance of $\sim 0.27 h^{-1} \text{ Mpc}$ at $z \sim 0.5$). It is then impossible, in any given observation, to obtain spectra of all galaxies with neighbours closer than this angular distance. The problem is alleviated in regions covered by multiple exposures, but it is in general not possible to observe all objects in crowded regions.

In this work we correct for the impact of fibre collisions on our clustering measurements by applying the correction presented in Guo et al. (2011). Using this method the total galaxy sample D is divided into two subsamples, dubbed as D_1 and D_2 . These are constructed following the targeting algorithm of the catalogue in a way that guarantees that group D_1 is not affected by fibre collisions, while D_2 contains all collided galaxies. Any clustering measurement of the combined sample can be obtained as a combination of the contributions from these two groups. Based on tests on mock galaxy catalogues, Guo et al. (2011) showed that the application of this method can accurately recover the projected and redshift-space correlation functions on scales both below and above the fibre collision scale, providing a substantial improvement over the commonly used nearest neighbour and angular correction methods.

We constructed random catalogues for subsamples D_1 and D_2 for the northern and southern hemispheres with 40 times more objects than the real data following their respec-

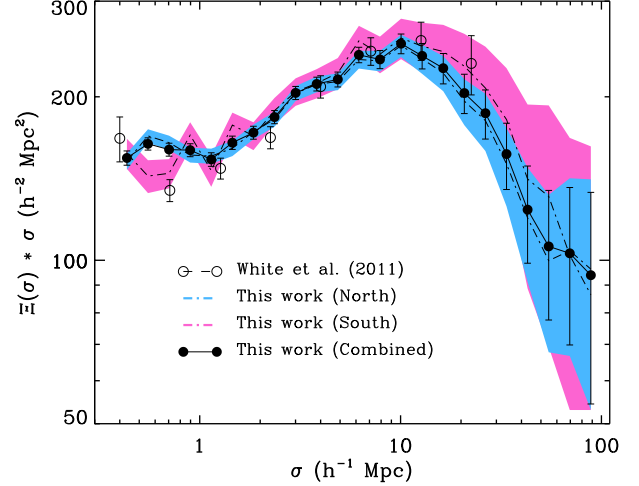


Figure 3. Projected correlation function times the projected distance for the DR9 BOSS-CMASS galaxy sample in the redshift range $0.43 < z < 0.7$. The blue and red shaded areas correspond to the north and south subsamples and give an estimate of their standard deviation. The dot-dashed lines display their mean value. The result of combining both subsamples is shown as filled circles. Standard deviation for the projected correlations of all samples are estimated using an ensemble of 600 mock catalogs (see Section 2.2). For comparison the projected correlation inferred from the first semester of the BOSS-CMASS data is also shown (open circles; White et al. 2011).

tive angular completenesses. The redshifts of these random points were generated in order to follow the distributions of the real samples, which were obtained by a smoothing spline interpolation of the observed redshift distributions.

With the increasing size of current galaxy surveys, and the corresponding improvement on the statistical uncertainties, the contribution of systematic errors to the total error budget of any clustering statistic becomes increasingly important. Due to its large volume and high number density BOSS is perhaps one of the best examples of this. Ross et al. (2012) present a detailed analysis of the systematic effects that could potentially affect any clustering measurement based on the CMASS sample and show that, besides redshift failures and fibre collisions, other important systematics must be considered in order to obtain accurate clustering measurements. The main result from this analysis is that these systematics can be corrected for by applying a set of weights, w_{sys} , which depend on both, the galaxy properties and their positions in the sky. We consider these weights in all our clustering measurements.

Finally, we also include a set of weights to reduce the variance of the estimator that are given by

$$w = (1 + n(z)J_w)^{-1}, \quad (4)$$

where $n(z)$ is the mean galaxy density at redshift z and J_w is a free parameter. Hamilton (1993) showed that setting $J_w = 4\pi J_3(s)$, where $J_3(s) = \int_0^s \xi(s')s'^2 ds'$, minimizes the variance on the measured correlation function for the given scale s . Here we follow the standard practice and use a scale-independent value of $J_w = 2 \times 10^4$.

Fig. 3 shows the projected correlation functions $\Xi(\sigma)$

times the projected distance of the north, south and combined CMASS samples. The combined sample gives a similar outcome to that of the north as a result of the higher statistics in the latter. For comparison the projected correlation inferred from a CMASS sample corresponding to the first semester of the BOSS observations is also shown (open circles; White et al. 2011). Besides the increase in the sample size and the volume probed, there are differences at small and large scales which are probably due to the different corrections for fibre collisions and the use of the weights to correct for the systematics affecting the galaxy density field.

Although the projected correlation functions of the northern and southern subsamples agree within their respective uncertainties, they show some intriguing differences. At scales in the range $\sim 20\text{--}50 h^{-1} \text{ Mpc}$ the amplitude of $\Xi(\sigma)$ in the south is higher than that of the north. Similarly, the measurements of $\xi(s)$ show the same behaviour. However, in this case, the agreement of the mean values is somewhat closer (see section 4).

To estimate covariance matrices for these clustering measures, we use a set of 600 mock catalogs designed to follow the same geometry and redshift distribution of the CMASS sample while mimicking their clustering properties at large scales (Manera et al. 2012). These mocks are inspired by the *PTHalos* method of Scoccimarro & Sheth (2002), although there are some important differences. The resulting covariances are compatible with the results of N -body simulations. For a detailed description about these mocks and their comparison with N -body results (see Manera et al. 2012)¹.

3 CLUSTERING IN THE Λ CMD MODEL

3.1 The MultiDark simulation

The MultiDark run (MDR1) is an N -body cosmological simulation of the Λ CDM model that was done using the Adaptive-Refinement-Tree (ART) code (Kravtsov et al. 1997; Gottlöber & Klypin 2008). The simulation has $2048^3 \approx 8.6 \times 10^9$ dark matter particles in a box of $1 h^{-1} \text{ Gpc}$ on a side. The mass of the dark matter particle is $8.72 \times 10^9 h^{-1} M_\odot$. The cosmological parameters adopted in the simulation are consistent with the latest WMAP7 results (Jarosik et al. 2011) and with other cosmological probes (see Table 1 of Klypin et al. 2011). Hence, we adopt a matter density parameter $\Omega_M = 0.27$ and a dimensionless Hubble parameter $h = 0.7$. Initial conditions were set at the redshift $z_{\text{init}} = 65$ using a power spectrum characterized by a scalar spectral index $n_s = 0.95$ and normalized to $\sigma_8 = 0.82$ in the same way as done for the Bolshoi simulation (see Klypin et al. 2011, for a detailed description of this simulation). The ART code is designed in such a way that the physical resolution is nearly preserved over time with a value of $\sim 7 h^{-1} \text{ kpc}$ for the redshift range between $z = 0\text{--}8$. For further details on the ART code and MultiDark simulation see Prada et al. (2011) and references therein.

3.1.1 Halo finding

Dark matter haloes are identified in the simulation with a parallel version of the Bound-Density-Maxima (BDM) algorithm (Klypin & Holtzman 1997; Riebe et al. 2011). The BDM is a Spherical Overdensity (SO) code. It finds all density maxima in the distribution of particles using a top-hat filter with 20 particles. For each maximum the code estimates the radius within which the overdensity has a specified value. Among all overlapping density maxima the code finds the one having the deepest gravitational potential. The position of this maximum is the centre of a “distinct” halo, which is a halo whose centre is not inside the virial radius of a bigger one. Distinct haloes are also tracers of central galaxies. Self-bound haloes with more than 20 particles lying inside the virial radius of a distinct halo are classified as subhaloes. Subhalo identification is more subtle since it requires the removal of unbound particles and identification of fake satellites. See Riebe et al. (2011) for a more detailed description of the algorithm. The BDM halo finder was extensively tested and compared with other halo finders (Knebe et al. 2011; Behroozi et al. 2011). In Appendix A we show a comparison between the real-space correlation function for halo catalogs selected both with BDM and RockStar halo finders (see Fig A1). The BDM halo catalogs for the MDR1 simulation are publicly available at the MultiDark Database: <http://www.multidark.org>.

The size of a distinct halo can be defined by means of the spherical radius within which the average density is Δ times higher than the critical density of the Universe, $\rho_{\text{cr}}(z)$. As a consequence, the corresponding enclosed mass is given by

$$M_\Delta = \frac{4\pi}{3} \Delta \rho_{\text{cr}}(z) R_\Delta^3. \quad (5)$$

We use a threshold overdensity of $\Delta = 200$ that corresponds to values for halo mass and radius of M_{200} and R_{200} . BDM catalogs also provide virial masses and radius (M_{vir} and R_{vir}) defined using the standard overdensity $360 \rho_{\text{back}}(z)$ (background mean density).

One of the most important characteristics of a (sub)halo is its maximum circular velocity:

$$V_{\text{max}}^2 = \max \left[\frac{GM(< r)}{r} \right]. \quad (6)$$

There are several advantages of using V_{max} to characterize a halo as opposed to the “virial mass”. First, V_{max} does not have the ambiguity related with the definition of mass. Virial mass and radius vary depending on the overdensity threshold used. For the often-employed overdensity 200 and “virial” overdensity thresholds, the differences in definitions result in changes in the halo radius from one definition to another and, thus, in concentration, by a factor of 1.2–1.3, with the exact value dependent on the halo concentration. Second and more important, the maximum circular velocity V_{max} is a better quantity to characterize haloes when we relate them to the galaxies inside these haloes. For galaxy-size haloes the maximum circular velocity is defined at a radius of $\sim 40 \text{ kpc}$, i.e., closer to the sizes of luminous parts of galaxies than the much larger virial radius, which for the Milky-Way halo is $\sim 250 \text{ kpc}$ (Klypin et al. 2002).

¹ Mocks will be available in <http://www.marcmanera.net/mocks/>

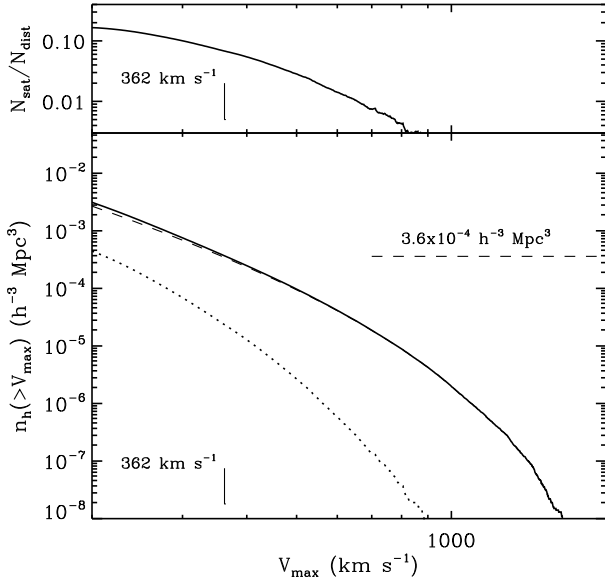


Figure 4. *Bottom panel:* The cumulative number density of distinct haloes (dashed line) and subhaloes (dotted line) in the MultiDark simulation at $z = 0.53$ as a function of maximum circular velocity. The cumulative number for all haloes is also shown as a solid line. *Top panel:* The cumulative subhalo fraction as a function of halo maximum circular velocity. As a reference we indicate in both panels the mean number density of the BOSS-CMASS galaxy sample and as vertical lines the corresponding maximum circular velocity threshold (V_{cut}) used in the HAM procedure.

3.2 Bridging the gap between galaxies and haloes

Once we have the maximum circular velocities for distinct haloes and subhaloes the implementation of the HAM prescription is simple. We start with a monotonic assignment. We count all haloes and subhaloes, which have maximum circular velocity V_{max} larger than V_{cut} , and gradually decrease the value of V_{cut} until the number density of (sub)haloes is equal to that of BOSS galaxies at $z \approx 0.5$.

The bottom panel of Fig. 4 shows the number density of (sub)haloes in the MultiDark simulation at $z = 0.53$. A number density close to that of the BOSS-CMASS sample corresponds to haloes and subhaloes with a maximum circular velocity above 362 km s^{-1} , which is larger than the completeness limit of the MultiDark simulation, i.e., $\sim 180 \text{ km s}^{-1}$. This means that haloes and subhaloes hosting BOSS-CMASS galaxies are well resolved. The top panel of Fig. 4 shows the cumulative subhalo fraction as a function of maximum circular velocity. For values of $V_{\text{max}} > 350 \text{ km s}^{-1}$ the subhalo fractions are typically less than 10%. We will return to this point again in Section 5.

3.2.1 Halo stochasticity

There are a number of arguments why there should be some degree of stochasticity in the stellar mass – circular velocity relation (e.g., Tasitsiomi et al. 2004; Behroozi et al. 2010; Trujillo-Gomez et al. 2011). In our case the stochasticity means that some haloes above the velocity cut host galaxies with stellar masses smaller than the corresponding stellar

mass cut of the BOSS sample and should not be included into the sample. Simultaneously, some smaller haloes may host galaxies with a larger stellar mass, and should be considered. Because the number density of galaxies is fixed by observations, the numbers of included and excluded haloes must be equal. Following Trujillo-Gomez et al. (2011) we implement this process using a Gaussian spread with an offset. If V_{cut} is the velocity cut in the monotonic assignment, then a (sub)halo is taken if its maximum circular velocity V_{max} satisfies the condition

$$V_{\text{max}} [1 + \mathcal{N}(0, \sigma)] - \Delta V > V_{\text{cut}}, \quad (7)$$

where $\mathcal{N}(0, \sigma)$ is a Gaussian random number with mean zero and *rms* σ . The offset ΔV is needed to compensate the larger influx of smaller haloes. We use $\sigma = 0.2$ and $\Delta V = 18 \text{ km s}^{-1}$, which is similar to the values adopted by Trujillo-Gomez et al. (2011). Note that the offset ΔV and the spread σ are not free parameters. The offset is just a normalization. The value of σ is defined by the spread of the observational Baryonic Tully-Fisher relation (or its equivalent for early-type galaxies), which has uncertainties (e.g., Trujillo-Gomez et al. 2011). The stochastic assignment has a very small effect on clustering for scales larger than $0.5 h^{-1} \text{ Mpc}$ decreasing the correlation functions no more than $\sim 8\%$.

3.2.2 Subhalo tidal stripping

In order to apply our HAM technique we use the maximum circular velocity at $z = 0$ as a proxy, which is a quantity that can be easily measured for haloes in our simulation. However, it is generally accepted that, for subhaloes, a better characteristic would be the peak value of the maximum circular velocity, V_{peak} , during subhalo evolution (e.g., Conroy et al. 2006; Trujillo-Gomez et al. 2011). The latter is motivated by the tidal stripping effect: once a halo falls into the potential well of a larger one some of its material can be stripped away, thus lowering the value of V_{max} . Since in real galaxies stars occupy the inner regions of subhaloes, where tidal forces are much weaker, their circular velocities should be, in general, less influenced by this effect.

We expect that the tidal stripping for BOSS-CMASS satellites, though present, not to be dominant, thus allowing us to use V_{max} for subhaloes as a reliable proxy for the HAM technique. In this case, satellites with masses of $\sim 10^{13} h^{-1} M_{\odot}$ are typically located at large distances from their central hosts, which can reach even larger mass values of $\sim 10^{14} - 10^{15} h^{-1} M_{\odot}$ (see Section 5).

To estimate the magnitude of the potential stripping effect in these systems we run a series of simple simulations. Using a direct-summation N -body code, we study the idealized case of a satellite orbiting its central host, where the latter is modeled as a static Navarro-Frenk-White (NFW) potential (Navarro et al. 1996). Initially, the satellite was set as a distribution of particles following an equilibrium NFW distribution with isotropic velocities. The mass per particle and force softening were set to $8 \times 10^7 h^{-1} M_{\odot}$ and $0.1 h^{-1} \text{ kpc}$ respectively. Particle mass decreases with decreasing distance to the central halo as a way to achieve a better mass resolution in the central regions. In order to

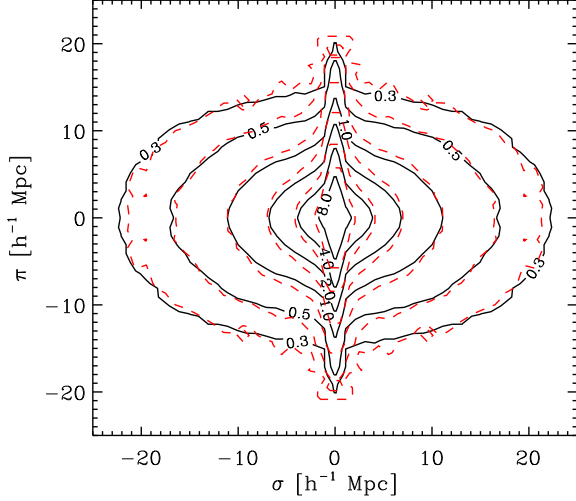


Figure 5. Contours of the two-dimensional correlation function $\xi(\sigma, \pi)$ estimated from the DR9 BOSS-CMASS north galaxy sample (dashed contours) at $0.4 < z < 0.7$ and for our MultiDark halo catalog constructed using the HAM technique at $z = 0.53$ (solid contours).

check for equilibrium stability and numerical effects we did a test run for an isolated satellite, i.e. without considering an external tidal field, finding that its maximum circular velocity was well preserved during the entire evolution of the system, which was set to 5 Gyr.

We study two different cases for a satellite of mass $M_{\text{sat}} = 10^{13} h^{-1} M_{\odot}$, alternatively assuming either $M_{\text{host}} = 10^{14} h^{-1} M_{\odot}$ or $M_{\text{host}} = 10^{15} h^{-1} M_{\odot}$ for the mass of the central host. Halo concentrations were selected to follow the results of Prada et al. (2011), thus they were taken to be $c_{\text{vir}} = 8.2, 6.9, 5.8$, in order of increasing halo mass. Stripping severely depends on the distance to the centre. For instance, for a central system with mass $M_{\text{host}} = 10^{14} h^{-1} M_{\odot}$, we find that a satellite with a pericentre (apocentre) of 100 (500) h^{-1} kpc loses around half of its maximum circular velocity in 5 Gyr. However, this is not typical of satellites in large galaxy clusters. We find that, for both host halo masses, satellites falling from the virial radius with apocentre-to-pericentre ratios of $\sim 4 : 1 - 3 : 1$, the tidal stripping is much less efficient, changing its maximum circular velocity only by 15–20% after 5 Gyr. The effect is much smaller after the first ~ 2 Gyr of evolution producing a variation less than 5%.

Since, in this work, the minimum studied physical scale is $\gtrsim 0.5$ Mpc, it is expected that most of the BOSS-CMASS satellites have spent most of their time at larger distances from their central haloes, where the impact of tidal forces is small. Thus, considering the relatively small change of the subhalo maximum circular velocities due to tidal stripping, we use V_{max} as a proxy for our HAM instead of V_{peak} . Interestingly, Watson et al. (2012), based on a subhalo evolution model applied to clustering measurements in the SDSS, suggest that tidal stripping of stars in luminous galaxies is much less efficient than in less luminous systems, which provides additional support to our choice.

3.3 Modeling BOSS-CMASS clustering

We use the MultiDark BDM catalogs constructed for the overdensity $360 \rho_{\text{back}}(z)$ to facilitate the comparison with the HOD modeling presented in White et al. (2011). However, as stated before, our results do not depend on halo mass definition since halo matching is done using the maximum circular velocity V_{max} of either distinct haloes or subhaloes. We use redshift $z = 0.53$, which is close to the peak value of the BOSS-CMASS sample (see Fig. 2).

To model the effect of galaxy peculiar velocities in the redshift measurements, we transform the coordinates of our (sub)haloes to redshift-space using $\mathbf{s} = \mathbf{x} + \mathbf{v} \cdot \hat{\mathbf{r}} / (aH)$, where \mathbf{x} and \mathbf{v} are their position and peculiar velocity vectors respectively, a is the scale factor and H is the Hubble constant. We compute the two-dimensional correlation function $\xi(\sigma, \pi)$ of our catalog counting the number of “galaxy” tracers in bins parallel (π) and perpendicular (σ) to the line-of-sight. When estimating the projected correlation function, we count all pairs along the parallel direction out to $\pi_{\text{max}} \sim 100 h^{-1}$ Mpc.

To estimate the cosmic variance we use a set of simulations from the Large Suite of Dark Matter Simulations (*LasDamas*; see <http://lss.phy.vanderbilt.edu/lasdamas/>). We use mock galaxy catalogs extracted from the *Carmen* boxes, which are 40 dark matter-only low-resolution runs done with 1120^3 particles in a periodic cube with $1 h^{-1}$ Gpc on a side. In this way, we can get a simple estimate of the expected *rms* deviations from our fiducial MultiDark result due to random fluctuations in the initial conditions of the universe. The dark matter density and scalar spectral index of the *Carmen* simulations display a difference of about 8% in comparison to the corresponding values of MultiDark. However, since here we only want to obtain an estimate for the magnitude of the cosmic variance, we consider this approach as good enough for this purpose.

As already mentioned in Section 2.2, to estimate the covariance matrices of observed correlation functions we use a set of 600 galaxy mocks designed to follow the same geometry and redshift distribution of the CMASS sample, while mimicking its clustering properties. Manera et al. (2012) show that the covariances for the correlation functions of *N*-body simulations are consistent with those resulting from the mocks. This means that it is safe to compare the cosmic variance of MultiDark (estimated from the *Carmen* set of simulations) with that resulting from the mock galaxy catalogs.

4 CLUSTERING OF GALAXIES IN THE BOSS-CMASS SAMPLE

The two-dimensional correlation function $\xi(\sigma, \pi)$ for the north subsample of BOSS-CMASS is presented in Fig. 5 for distances up to $\sim 20 h^{-1}$ Mpc (dashed contours). The Finger-Of-God elongation along the line-of-sight direction at small perpendicular separations, which is due to galaxy small-scale random velocities, is clearly seen. The flattening of contours at larger projected scales is due to the Kaiser effect caused by large-scale infall velocities (Kaiser 1987). Predictions for the clustering of galaxies obtained from the MultiDark cosmological simulation (solid contours) produce a

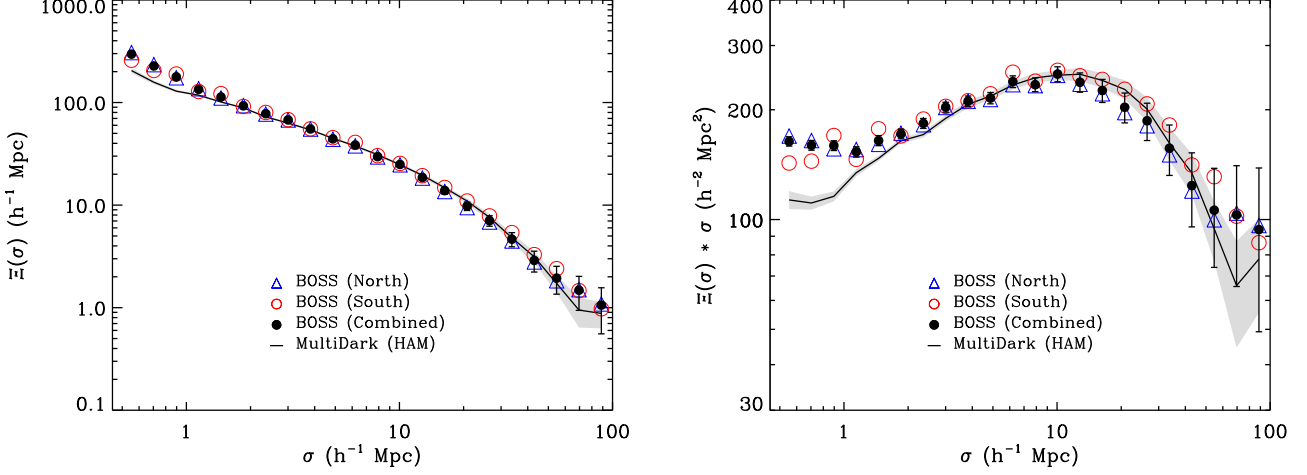


Figure 6. *Left panel:* Projected correlation function for the $0.4 < z < 0.7$ DR9 BOSS-CMASS north, south and Full galaxy samples (open blue triangles, open red circles and filled black circles respectively) and the MultiDark catalog selected with the HAM procedure at $z = 0.53$ (solid line). The shaded area for MultiDark gives an estimate of the cosmic variance. BOSS-CMASS error bars were estimated using an ensemble of 600 mock galaxies (see Section 2.2). For clarity, only error bars for the combined sample are shown. The corresponding ones for the north and south are a factor of about 1.13 and 2.15 times larger respectively. The transition between the one-halo and two-halo terms can be seen at $\sim 1 h^{-1} \text{ Mpc}$. Flattening of the signal at intermediate scales and bending at large scales are also evident features. *Right panel:* Detailed differences between the ΛCDM model and BOSS clustering are better seen when plotting the quantity $\Xi(\sigma) \sigma$.

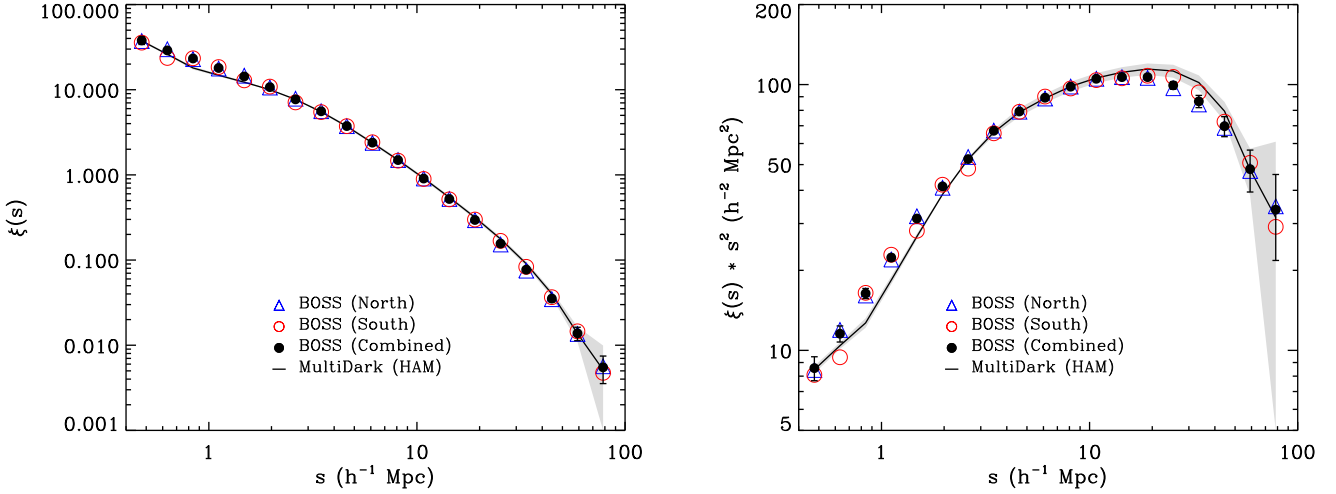


Figure 7. *Left panel:* Redshift-space correlation function for the $0.4 < z < 0.7$ DR9 BOSS-CMASS north, south and Full galaxy samples (open blue triangles, open red circles and filled black circles respectively) and the MultiDark catalog selected with the HAM procedure at $z = 0.53$ (solid line). Standard deviation for model and observations are shown in the same way as in Fig. 6. *Right panel:* Shown is the quantity $\xi(s) s^2$ which better reflects the differences between our ΛCDM model and BOSS clustering measures.

fair representation of the measured clustering in the CMASS sample. Nevertheless, there are some deviations. At small separations, $\sigma \lesssim 2 h^{-1} \text{ Mpc}$, observations show more clustering as compared with results from the simulation. The situation reverses at large scales, where our cosmological simulation predicts slightly stronger clustering.

These tendencies are clearly seen in the correlation functions presented in Figs. 6 and 7. The north, south and combined CMASS samples are shown together with the result of our simple HAM model. The shaded area for MultiDark

gives an estimate of the cosmic variance as computed from *LasDamas* suite of simulations. Again, the overall agreement at all scales is quite good showing a remarkable match with observations. However, as noted before, there are some noticeable discrepancies at small and intermediate scales. The detailed differences between the projected correlation function and MultiDark can be better seen in the right panel of Fig. 6, where differences in the correlations are amplified after multiplying by the corresponding projected distance. The disagreement at scales $\lesssim 1 h^{-1} \text{ Mpc}$ is perhaps

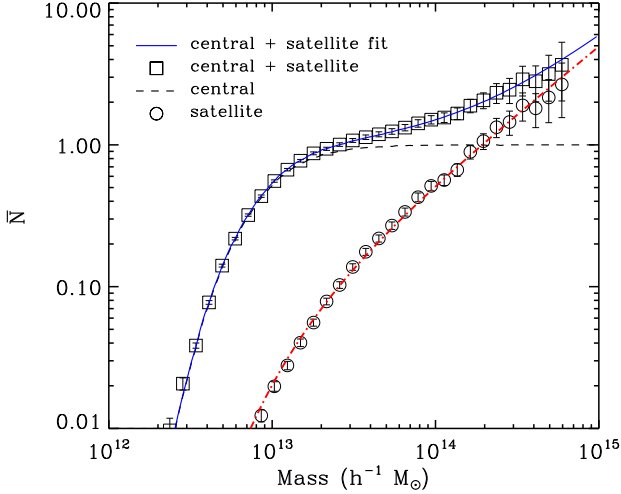


Figure 8. The mean occupancy of all haloes in our MultiDark sample used to match the BOSS-CMASS observations as a function of halo mass (open squares). Open circles and dashed line correspond to satellites and central haloes respectively. Error bars are calculated assuming Poisson statistics in the counting. The fit given by Eq. (8) is shown as a dot-dashed line.

related to the simple stochastic HAM adopted here. At large scales, starting from $\sim 20 h^{-1}$ Mpc, the theoretical estimates lie slightly above the observational estimates of the northern galaxy subsample (at $\sim 1\sigma$ level), which has about four times larger statistics than the corresponding southern sample.

The redshift-space clustering results, both for the CMASS sample and the Λ CDM model given by the MultiDark simulation, are shown in Fig. 7. As before, the shaded area represents cosmic variance estimates and differences between model and observations are better seen in the right panel. Peculiar velocities of galaxies inside virialized systems reduce the clustering signal thus lowering the slope of the correlation function at scales of $1\text{--}2 h^{-1}$ Mpc.

For scales in the range $\sim 0.6\text{--}1 h^{-1}$ Mpc our model underpredicts the observed values, as already showed for the case of the projected correlation function. The agreement between the simulation measurement and observed redshift-space correlation function is quite remarkable at scales $\gtrsim 1 h^{-1}$ Mpc. Differences are less than 2–3% for a wide range of distances ranging from $2 h^{-1}$ Mpc to $20 h^{-1}$ Mpc. At $20\text{--}40 h^{-1}$ Mpc the MultiDark results overpredict the observed clustering by about $\sim 10\%$. Statistically the differences are significant: the effect is about 3σ at $\sim 30 h^{-1}$ Mpc (e.g., at $s = 33.5 h^{-1}$ Mpc the redshift-space correlation function for the combined CMASS sample and MultiDark give $\xi_{N+S}(s) = 0.077 \pm 0.004$ and $\xi_{MD}(s) = 0.091 \pm 0.003$, respectively). The small differences between the N -body results and observations may be alleviated if we use a more sophisticated HAM procedure including, for instance, light-cone effects and a match to the stellar mass distribution at these redshifts. Nevertheless, the high level of agreement found between data and observations using the simple HAM procedure adopted here is a striking result.

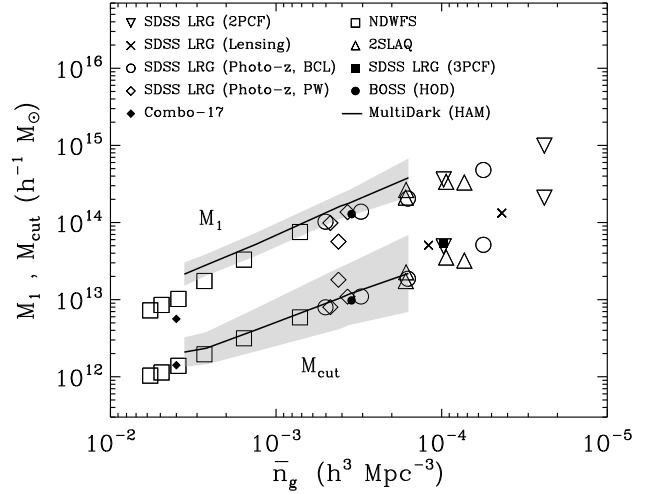


Figure 9. MultiDark HOD parameters, M_{cut} and M_1 , as a function of number density (solid line) using the simple HAM prescription at $z = 0.53$. We compare our results with a variety of intermediate redshift massive galaxy samples. The data are taken from Phleps et al. (2006), Mandelbaum et al. (2006), Kulkarni et al. (2007), Blake et al. (2008), Brown et al. (2008), Padmanabhan et al. (2009), Wake et al. (2008) and Zheng et al. (2009). Filled circles show results from White et al. (2011) HOD's analysis of early BOSS data (see text).

5 THE MEAN HALO OCCUPANCY OF BOSS-CMASS GALAXIES

Our analysis allows us also to study the halo occupation distribution and the satellite fraction of BOSS-CMASS galaxies at $z \sim 0.5$. The main advantage of the MultiDark simulation is that it has sufficient resolution to resolve satellites around central distinct haloes. The satellite distribution around massive haloes can be directly studied from the resulting halo catalogs. As shown previously in the top panel of Fig. 4, the fraction of satellites for haloes with a number density close to that of the CMASS sample is less than 10%. In particular, for haloes having $V_{\text{max}} \geq 362 \text{ km s}^{-1}$, which corresponds to a number density of $3.6 \times 10^{-4} h^{-3} \text{ Mpc}^3$, the resulting satellite fraction is 6.8%. The HOD modeling by White et al. (2011), using the first semester of BOSS data, reported a satellite fraction $(10 \pm 2)\%$ which is reduced to $(7 \pm 2)\%$ when they ignore in their fit to the correlation function the very small scales affected by fibre collisions. Note that our HAM procedure is non-parametric and provides satellite fractions consistent with our Λ CDM cosmological simulation. Yet, the halo-occupancy distribution and satellite fractions from HOD modeling are obtained from a fit to the empirical correlation function.

Fig. 8 shows the mean occupancy of haloes for the BOSS-CMASS sample as obtained from the MultiDark halo abundance-matching scheme. The dashed line and open circles are the contributions of distinct haloes and subhaloes respectively. Open squares correspond to the total occupancy of haloes, including both central and satellite galaxies from our halo catalog. Distinct haloes display a sharp transition around $M_{\text{vir}} \gtrsim 10^{13} h^{-1} M_{\odot}$. The mean number of satellite

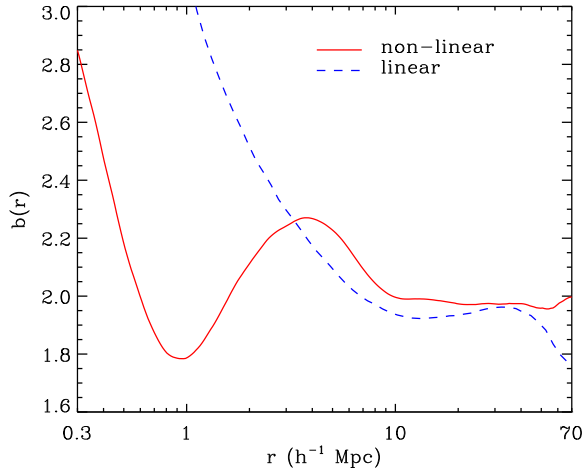


Figure 10. Scale-dependent galaxy bias at $z = 0.53$ predicted for BOSS-CMASS galaxies using the MultiDark simulation. The solid curve shows the bias relative to the dark matter (Eq. (9)). The bias relative to the linear-theory predictions is shown as a dashed line.

galaxies as a function of halo mass can be modeled with the following expression (e.g., Wetzel & White 2010)

$$\bar{N}_{\text{sat}}(M) = \left(\frac{M}{M_1}\right)^\alpha e^{-M_{\text{cut}}/M_1}, \quad (8)$$

where $\log M_{\text{cut}} = 13.07 \pm 0.40$, $\log M_1 = 14.25 \pm 0.17$ and $\alpha = 0.94 \pm 0.42$ (dot-dashed line). Here, M_1 is the halo mass which hosts, approximately, one satellite and M_{cut} governs the strength of the transition between systems with and without satellite systems. For high halo masses, fluctuations in the determination of the satellite occupancy arise because we are dealing with small number statistics as a result of the fixed volume of the simulation. The solid line in Fig. 8 shows the total mean halo occupancy but using in this case the best fit model for the satellite distribution in order to extrapolate the result towards higher masses.

In Fig. 9 we compare the HOD parameters, M_{cut} and M_1 , obtained from MultiDark at $z = 0.53$ as a function of number density (solid lines) following our HAM scheme. We also show estimates for a variety of intermediate redshift massive galaxy samples from the literature, including the HOD results from White et al. (2011) for the early BOSS data sample. This compilation of different datasets has been kindly provided by M. White. Error bars on the individual points are typically ~ 0.1 dex, as represented by the size of the symbols. The agreement between the MultiDark HAM predictions and data from different surveys is remarkable if one considers the differences in sample selection, redshift range and HOD methods. Our estimates for the HOD parameters of the BOSS-CMASS sample yield consistent values with those of White et al.’s HOD analysis which are contained within our error bars. Nevertheless, it is worth noting here that White et al. (2011) did not consider weights in the estimation of the correlation function used, which could have an impact on the derived parameters, and that our approach relies completely on our halo catalog.

6 POWER SPECTRUM AND BIASES

In this section we focus on the abundance-matched halo catalog to the BOSS-CMASS galaxy sample, presenting further predictions from the MultiDark simulation that can be tested with future observations. Using the resulting halo sample and dark matter particles from the simulation we can estimate the bias of the halo population with respect to the underlying mass distribution as follows

$$b(r) \equiv \sqrt{\frac{\xi_h(r)}{\xi_m(r)}} \quad (9)$$

where $\xi_h(r)$ and $\xi_m(r)$ are the real space correlation functions for the MultiDark haloes and dark matter in the volume at the redshift of interest. This bias is shown in Fig. 10 as a function of spatial scale (solid line). The resulting bias is $b \sim 2$ at the transition scale of $\sim 1 h^{-1}$ Mpc and, as expected, increases strongly for smaller scales where galaxies are more strongly clustered with respect to the dark matter. For the remaining scales we can constrain the bias factor to be in the range $b \approx 1.8$ – 2.2 . Interestingly, this result is in very good agreement with the findings of Ho et al. (2012). These authors found a galaxy bias of $b = 1.98 \pm 0.05$ in the redshift range $z = 0.50$ – 0.55 by studying the angular clustering of the photometric CMASS sample. The bump-like feature between ~ 1 – $10 h^{-1}$ Mpc is related to the transition between the one- and two-halo terms in the correlation function, while for larger scales the bias factor tends to decrease. The linear bias estimation is shown as a dashed line, where the linear matter correlation function is used instead. As expected, the linear bias at small scales differs strongly from the non-linear result while approaching more similar values at larger scales.

We have a number of goals with the analysis of the power spectrum and biases: (1) We want to present accurate approximations for the numerical results, which can be used for comparison of observational results with predictions of the cosmological model used in our simulations. It is more convenient to use these approximations instead of having to deal with raw simulations. (2) The high quality of our results allows us to study effects which are difficult to measure with low-resolution simulations.

One should clearly understand the role of the standard Λ CDM model with the particular set of cosmological parameters used for our simulations. Our results show that, once we match the abundance of haloes, the model reasonably reproduces a wide range of scales of the observed projected and redshift-space correlation functions. In principle, one can invert the correlation function to obtain the power spectrum. However, in practice, a model-independent inversion is a technically complicated process. This is why we chose a different approach: we use the power spectrum of haloes in the model as a proxy of the power spectrum of galaxies in BOSS.

We use two other sets of simulations in addition to MultiDark. The first one is the already mentioned *Carmen* series of 40 simulations of the *LasDamas* suite of simulations that allow us to estimate the effect of cosmic variance. These mock galaxy catalogs are produced with an HOD model with parameters aimed at fitting the respective SDSS galaxy samples (McBride et al. 2009). Note that,

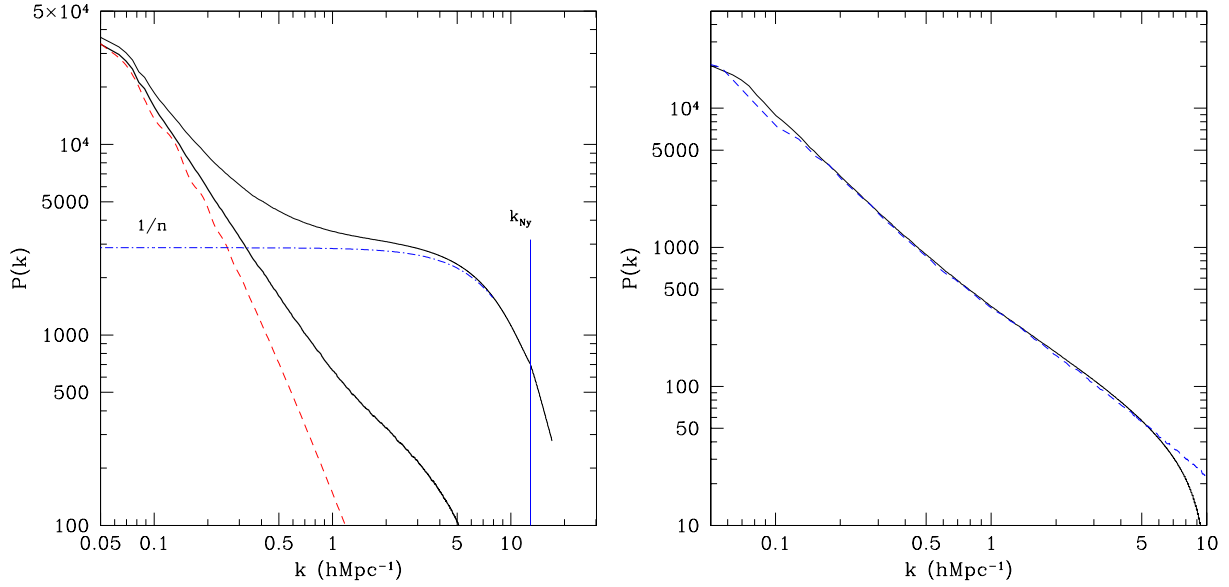


Figure 11. *Left panel:* Recovering the power spectrum: shot-noise and density assignment corrections. The top solid thin curve shows the “raw” estimate of the power spectrum at $z = 0.53$ for haloes and subhaloes with circular velocities larger than $V_{\max} > 362 \text{ km s}^{-1}$ corresponding to a number density close to that of galaxies in the BOSS sample $n = 3.6 \times 10^{-4} h^3 \text{ Mpc}^{-3}$. The dot-dashed line is the combined correction in Eq. (10) due to the shot-noise and the density assignment. The vertical line shows the Nyquist frequency. The thick solid line is the recovered power spectrum. The dashed line shows the linear power spectrum of dark matter density perturbations scaled up to match the amplitude of the recovered power spectrum at long waves. *Right panel:* Comparison between the recovered power spectra for haloes+subhaloes with $V_{\max} > 200 \text{ km s}^{-1}$ in the MultiDark (solid line) and the Bolshoi (dashed line) simulations at $z = 0$. Deviations at $k < 0.1 h \text{ Mpc}^{-1}$ are due to cosmic variance. The deviations at $k > 5 h \text{ Mpc}^{-1}$ are due to density assignment effects in the MultiDark simulation. However, for wave-numbers in the range $0.2 h \text{ Mpc}^{-1} < k < 5 h \text{ Mpc}^{-1}$ the resulting power spectra are not affected by cosmic variance and resolution and the agreement between simulations is excellent, with deviations less than just few percent.

as before, we use only relative model-to-model deviations in the *Carmen* simulations: error bars in our results are obtained in this way. Secondly, we also use results of the Bolshoi simulation (Klypin et al. 2011). This simulation has a factor of ~ 5 better mass and force resolution, but it was performed for a smaller simulation box ($250 h^{-1} \text{ Mpc}$ on a side). There is an overlap between the MultiDark and Bolshoi simulations: the simulation volume of Bolshoi is large enough to study haloes (and subhaloes) with circular velocities of $\sim 200 \text{ km s}^{-1}$. At the same time, these (sub)haloes are reasonably well resolved in the MultiDark simulation having more than 100 particles. Comparison of MultiDark and Bolshoi power spectra for these haloes allows us to look for biases at scales $k > 0.1 h \text{ Mpc}^{-1}$.

To estimate power spectra, we use a large density mesh of 4096^3 cells and then we apply the standard FFT method. The Cloud-In-Cell density assignment scheme is used to calculate the density fields from the coordinates of haloes in the simulations. However, before the power spectra can be reliably used two corrections should be applied: a correction due to the density assignment (Jing 2005) and the usual shot-noise correction. If the number density of objects is $n = N/L^3$ and the Nyquist wave-number is $k_{\text{Ny}} = \pi N_{\text{grid}}/L$, then the corrected power spectrum is given by

$$P(k) = P_{\text{raw}}(k) - \frac{1}{n} \left[1 - \frac{2}{3} \sin^2 \left(\frac{\pi k}{2k_{\text{Ny}}} \right) \right], \quad (10)$$

where L is the length of the computational box and $N_{\text{grid}} = 4096$. This approximation is known to work well for $k <$

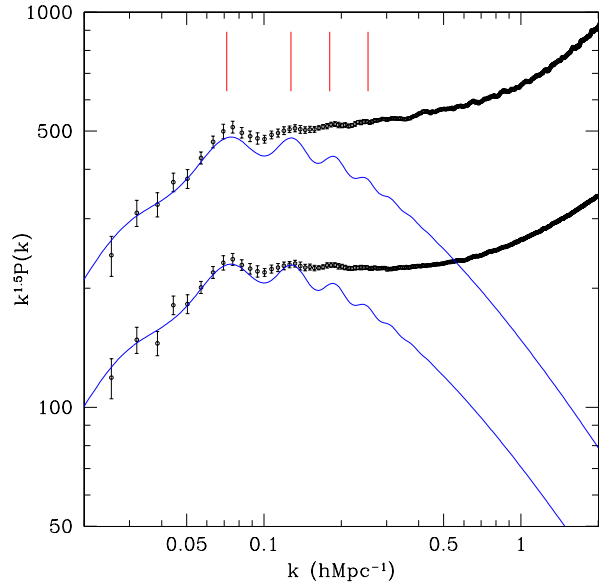


Figure 12. Power spectra (multiplied by $k^{1.5}$) of dark matter haloes in real space (open circles with error bars) for haloes with $V_{\max} > 362 \text{ km s}^{-1}$ (top) and $V_{\max} > 180 \text{ km s}^{-1}$ (bottom). Solid curves show the linear power spectra scaled to match the amplitude of fluctuations at long waves. The four vertical lines indicate the positions of maxima due to BAO. The BAO peaks in the linear spectrum give rise to peaks in the power spectrum of haloes.

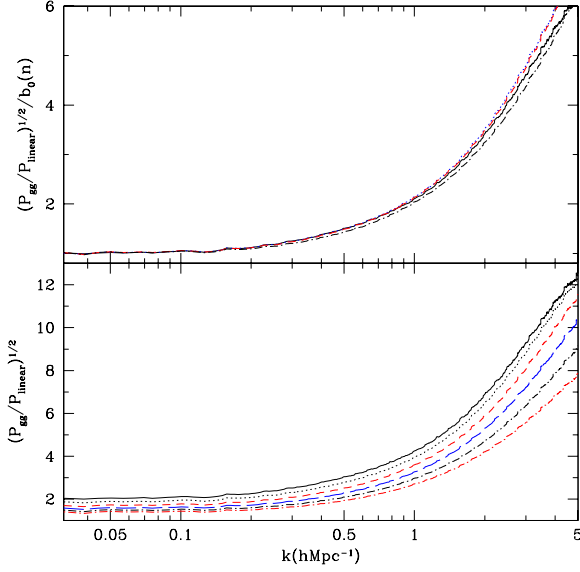


Figure 13. *Bottom panel:* Real-space bias factor $b(k) = (P_{\text{gg}}/P_{\text{linear}})^{1/2}$ for haloes with circular velocities $V_{\text{max}} = 180, 200, 220, 250, 300$ and 362 km s^{-1} (from bottom to top). *Top panel:* Bias factor for different haloes normalized to unity at long-waves. The bias factor $b(k)$ depends on the circular velocity V_{max} and on wave-number k in a rather complicated way. There are small depressions in the bias factor at peaks of BAOs. When normalized to the long-wave value b_0 , the bias factor is slightly smaller for less massive haloes. However, the main effect is the overall shift b_0 .

$0.7k_{\text{Ny}}$ (Jing 2005; Cui et al. 2008). However, to remain on safe ground we decided to limit our analysis to $k < 0.4k_{\text{Ny}} = 5 h \text{ Mpc}^{-1}$. The left panel of Fig. 11 illustrates the procedure of shot-noise and density corrections using a halo sample with $V_{\text{max}} > 362 \text{ km s}^{-1}$ extracted from the MultiDark simulation at $z = 0.53$.

In the right panel of Fig. 11 we compare results of the MultiDark and Bolshoi simulations. Just as one may expect, there are some deviations at long waves due to the cosmic variance: the Bolshoi box of $250 h^{-1} \text{ Mpc}$ is too small to accurately probe these waves. There are also deviations at short waves that correspond to $k > 7 h \text{ Mpc}^{-1}$ that are mainly due to the difference in density assignment between both simulations. For the Bolshoi simulation, the adopted mesh sets a minimum physical scale four times higher in frequency in comparison to MultiDark. However, for wave-numbers in the range $0.2 h \text{ Mpc}^{-1} < k < 5 h \text{ Mpc}^{-1}$ the agreement between the simulations is remarkably good. This agreement is especially important for short waves, where both resolution and shot-noise could have corrupted the results. However, since this has not happened, it indicates that the obtained power spectrum for MultiDark can be trusted up to, at least, $k = 5 h \text{ Mpc}^{-1}$.

Fig. 12 shows power spectra of haloes with circular velocity cuts $V_{\text{max}} > 362 \text{ km s}^{-1}$ (top curves) and $V_{\text{max}} > 180 \text{ km s}^{-1}$ (bottom curves). To highlight BAO features, we actually plot the power spectra of the halo distribution multiplied by $k^{1.5}$. As a result, the first four peaks in the spectra are clearly seen in the plot. However, they are somewhat smeared out by the non-linear evolution. As expected,

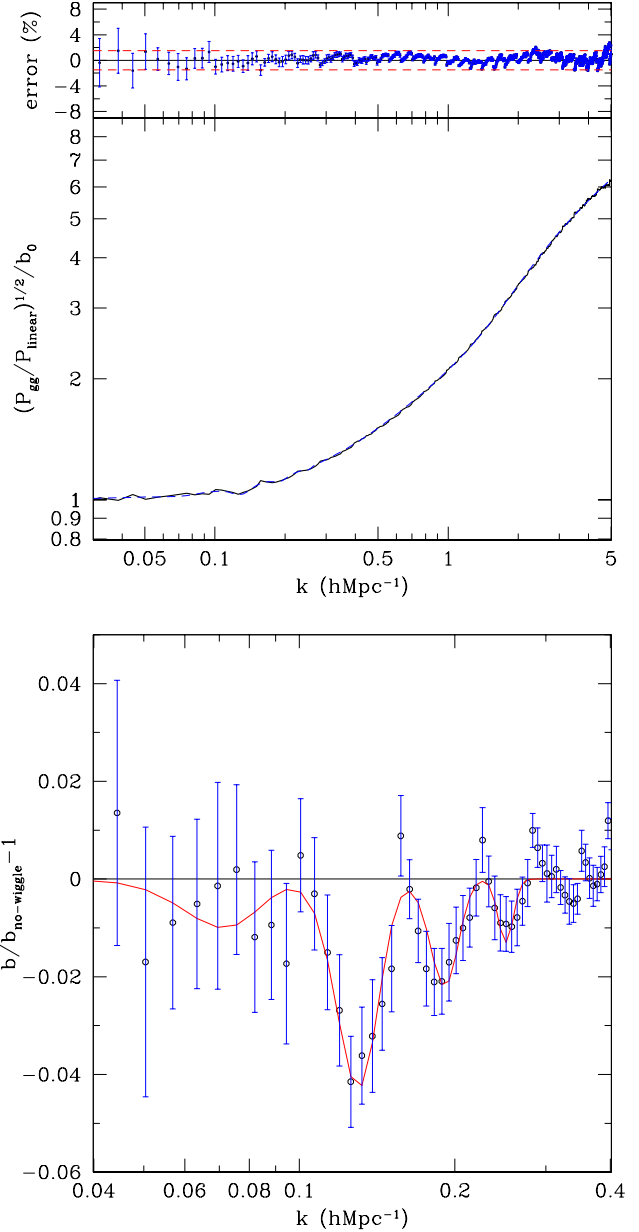


Figure 14. Real-space bias factor for haloes with circular velocities larger than $V_{\text{max}} = 362 \text{ km s}^{-1}$. *Top panel:* The bias factor normalized to the long-wave value b_0 (bottom panel, solid line) is compared with the analytical approximation given by Eq. (15) (dashed line). The top panel displays the relative error in percentages of the analytical approximation (filled circles). Error bars show the *rms* fluctuations due to the cosmic variance. *Bottom panel:* Deviations of the bias from the “de-wiggled” component of the bias factor given by Eq. (15). Open circles show the relative deviations $b(k)/b_{\text{no-wiggle}} - 1$ for each wave-number. The solid line is an analytical model for the residuals: the sum of exponential terms in Eq. (15). Error bars show the *rms* fluctuations due to cosmic variance.

the smearing increases for larger wave-numbers where the non-linearity is more important.

In what follows, we define the bias factor by

$$b(k, V_{\max}) \equiv \left[\frac{P_{\text{gg}}(k, V_{\max})}{P_{\text{linear}}(k)} \right]^{1/2}, \quad (11)$$

where $P_{\text{linear}}(k)$ is the linear power spectrum of the dark matter and $P_{\text{gg}}(k, V_{\max})$ is the power spectrum of haloes and subhaloes with circular velocities larger than V_{\max} . In order to distinguish the latter from the often used non-linear dark matter power spectrum or from the power spectrum of distinct haloes, we use subscript “gg” to indicate that our results mimic galaxies.

We start our analysis with the long-wave normalization of the bias parameter for different velocity cuts and, thus, for different number-densities of our “galaxies”. The bottom panel in Fig. 13 shows $b(k, V_{\max})$ for different velocities. At all wave-numbers the bias $b(k, V_{\max})$ increases with increasing V_{\max} . The top panel shows that when normalized to the long-wave value, $b_0(V_{\max})$, the bias factor is nearly the same. However, there is some residual dependence on V_{\max} , i.e., the deviations of the bias from one velocity cut to another can be as large as 15% and this should be taken into account if an accurate fit is needed. An approximation for the real-space long-wave bias factor as a function of the average number density of dark matter haloes $n(> V_{\max})$ or V_{\max} is presented below:

$$b_0(n) = 1 + 0.57 \log_{10} \left(\frac{2.05 \times 10^{-2} h^3 \text{Mpc}^{-3}}{n} \right), \quad (12)$$

$$b_0(V_{\max}) = 1 + \left(\frac{V_{\max}}{361 \text{ km s}^{-1}} \right)^{4/3}. \quad (13)$$

We now focus our analysis on the bias factor of haloes with $V_{\max} > 362 \text{ km s}^{-1}$ at $z = 0.53$, whose abundance $n = 3.6 \times 10^{-4} h^3 \text{Mpc}^{-3}$ matches that of BOSS galaxies. The top panel of Fig. 14 shows the bias factor of these haloes normalized to the value $b_0(V_{\max}) = 2.01$ found at long waves. Overall the bias factor is nearly flat at long waves and monotonically increases to short waves. The following approximation for the smooth component of the real-space bias factor gives an accuracy better than 4%:

$$b(k, V_{\max}) = b_0(V_{\max}) [1 + \log_{10}(1 + 3k^{1.8} + 5.8k^3)], \quad (14)$$

where the wave-number k is in units of $h \text{Mpc}^{-1}$. However, this approximation misses an important effect of non-linearities, the damping of the BAO. The coupling between different Fourier modes washes out the acoustic oscillations, erasing the higher harmonic peaks (Meiksin et al. 1999; Eisenstein et al. 2007b; Angulo et al. 2005, 2008; Sánchez et al. 2008; Montesano et al. 2010). In recent years, there has been substantial progress in the theoretical understanding of non-linear distortions in the BAO signal, which can now be accurately modelled (see e.g., Crocce & Scoccimarro 2006, 2008; Matsubara 2008a,b; Taruya et al. 2009), and even partially corrected for (Eisenstein et al. 2007a; Seo et al. 2010). As the bias factor in Eq. (11) is defined with respect to the extrapolated linear theory power spectrum, this damping leads to small wiggles in $b(k)$ with an amplitude at the 2–4% level, that can be better seen in the bottom panel of Fig. 14.

Table 1. Parameters for the approximation of the real-space bias factor given by Eq. (15).

BAO peak	k ($h \text{Mpc}^{-1}$)	α_i	σ_i
1	0.071	0.010	0.017
2	0.130	0.043	0.017
3	0.191	0.022	0.017
4	0.251	0.013	0.012

The accuracy of the fitting of $b(k, V_{\max})$ for $V_{\max} > 362 \text{ km s}^{-1}$ can be improved by including extra terms in the expansion and by adding the four main BAO peaks as follows

$$b(k, V_{\max}) = b_0(V_{\max}) \times [1 + \log_{10}(1 + 4.0k^{1.8} + 3.1k^3 + 1.0k^{4.5})] \times \prod_i \left[1 - \alpha_i \exp \left(-\frac{(k - k_i)^2}{\sigma_i^2} \right) \right]. \quad (15)$$

Here each BAO peak is approximated as a small suppression of the bias factor given by the last term of the equation, k_i is the wave-number of the peak and $\alpha_i \approx 0.01$ –0.05 and $\sigma_i \approx 0.01$ –0.02 are free parameters. The typical errors given by this approximation are smaller than 2% (see the top panel of Fig. 14). The values of the parameters used in the approximation can be seen in Table 1.

Using Eq. (13) and the bias factor $b(k)$ for the velocity cut $V_{\max} = 362 \text{ km s}^{-1}$, we develop corrections to the bias factor for different values of V_{\max} . In this way, we find the following set of equations that yield an accuracy better than 4% for the range of velocities within $V_{\max} = 180$ –370 km s^{-1} :

$$b(k, V_{\max}) = b_0(V_{\max}) \times [1 + \log_{10}(1 + 4.0k^{1.8} + 3.1k^3 + 1.0k^{4.5})] \times \prod_i \left[1 - \alpha_i \exp \left(-\frac{(k - k_i)^2}{\sigma_i^2} \right) \right] \times [1 - \beta_0 (1 - e^{-k^2/0.22^2}) + \beta_1 k - \beta_2 k^3] \quad (16)$$

where the parameters $\beta_{0,1,2}$ depend only on V_{\max}

$$\begin{aligned} \beta_0 &= \left(\frac{66.6 \text{ km s}^{-1}}{V_{\max}} \right)^3, \\ \beta_1 &= 2.18 \times 10^{-2} \left[1 - \left(\frac{205.8 \text{ km s}^{-1}}{V_{\max}} \right)^{103/14} \right], \\ \beta_2 &= 1.64 \times 10^{-2} \left[1 - \left(\frac{266.5 \text{ km s}^{-1}}{V_{\max}} \right)^{1/6} \right]. \end{aligned} \quad (17)$$

7 CONCLUSIONS

We presented an analysis of the clustering of 282,068 galaxies in the DR9 sample of BOSS data for a wide range of scales, ranging from $\sim 500 h^{-1} \text{kpc}$ to $\sim 90 h^{-1} \text{Mpc}$. We separately studied the clustering in the northern and southern hemispheres, as well as for the full sky sample. We measured the two-dimensional, projected and redshift-space correlation functions and compare the results with those ob-

tained from a large cosmological simulation with $1 h^{-1}$ Gpc on a side at a redshift of $z = 0.53$. The cosmological parameters adopted in the simulation are consistent with the latest WMAP7 results and several other probes. Our simulation, also known as MultiDark, is able to resolve the relevant subhalo masses needed to compare with the observed satellite population. To bridge the gap between galaxies and dark matter haloes we use a simple HAM technique applied to the BOSS-CMASS sample. Our main results can be summarized as follows:

- There is a 10–20% asymmetry in the projected and redshift-space correlation functions between the north and south subsamples at $\gtrsim 20 h^{-1}$ Mpc scales, which is better seen in the case of the projected correlation function. However, for both subsamples, the mean values agree with each other within a $\sim 1\sigma$ level of uncertainty.

- As compared with the first-semester of BOSS results presented by White et al. (2011), we find a small increase in power in the projected correlation function at scales smaller than $\sim 1 h^{-1}$ Mpc due to the improved treatment of fibre collisions and new corrections for systematics. However, the correlation functions (projected and redshift-space) decline by 10–20% at $10\text{--}30 h^{-1}$ Mpc scales in comparison with our HAM model. This is most noticeable for the north subsample which has about four times larger statistics than its southern counterpart. The comparison with the south subsample yields more consistent results with MultiDark at all scales, both in the projected and redshift-space correlations.

- Our N -body predictions for the clustering of galaxies give a fair representation of the measured clustering in the CMASS sample for a wide range of scales. The more consistent results between the north and south subsamples for the redshift-space correlation function show a remarkable agreement with theory: the differences are of the order of $\sim 2\%$ on scales ranging from $2 h^{-1}$ Mpc up to $20 h^{-1}$ Mpc. This result is more impressive when considering the fact that our simple HAM scheme does not include any free parameter. At larger distances, however, we find some deviations when comparing with the north subsample. For scales in the range of $20\text{--}40 h^{-1}$ Mpc the theoretical redshift-space correlation function is above the observations by $\sim 10\%$. Statistically, this difference is important – e.g., it represents a $\sim 3\sigma$ deviation at $\sim 30 h^{-1}$ Mpc. Future data and a more sophisticated theoretical modeling may help to clarify the situation.

- The distribution of (sub)haloes as a function of halo mass, as measured from our abundance-matched halo catalog, points towards a BOSS-CMASS galaxy population inhabiting haloes of mass $M \gtrsim 10^{13} h^{-1} M_{\odot}$, with $\sim 7\%$ of them being satellites orbiting centrals with $M \gtrsim 10^{14} h^{-1} M_{\odot}$. We also derived values for the HOD parameters of the sample using our simulation: $\log M_{\text{cut}} = 13.07 \pm 0.40$ and $\log M_1 = 14.25 \pm 0.17$.

- The scale-dependent galaxy bias of BOSS galaxies is likely to be $b \simeq 2$ at scales $\gtrsim 10 h^{-1}$ Mpc (see Eq. (9)). Furthermore, using our simulation, we also compute a large-scale bias (defined as the ratio between the

abundance-matched galaxy catalog and the extrapolated linear matter power spectra; see Eq. (11)) and found that it depends on the galaxy maximum circular velocity as $b(V_{\text{max}}) = 1 + (V_{\text{max}}/361 \text{ km s}^{-1})^{4/3}$, or on the galaxy number density as $b(n_g) = 0.0377 - 0.57 \log_{10}(n_g/h^3 \text{ Mpc}^{-3})$. These approximations can be used to compare observational results with predictions for the cosmology adopted in our simulation.

- The large-scale galaxy bias, defined using Eq. (11), has $\sim 2\text{--}4\%$ dips at the positions of BAO peaks in the spectrum of fluctuations that are due to shifts caused by non-linear effects. In this case, we also provide very accurate fits of the bias as a function of maximum circular velocity of galaxies that can also be used to recover the non-linear galaxy power spectrum in terms of the extrapolated linear density field of matter.

ACKNOWLEDGMENTS

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, University of Florida, the French Participation Group, the German Participation Group, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

The MultiDark Database used in this paper and the web application providing online access to it were constructed as part of the activities of the German Astrophysical Virtual Observatory as a result of the collaboration between the Leibniz-Institute for Astrophysics Potsdam (AIP) and the Spanish MultiDark Consolider Project CSD2009-00064. The Bolshoi and MultiDark simulations were run on the NASA's Pleiades supercomputer at the NASA Ames Research Center.

S.E.N. and F.P. acknowledges support from the Spanish MICINN's Consolider grant MultiDark CSD2009-00064. S.E.N. also acknowledges support by the Deutsche Forschungsgemeinschaft under the grant MU1020 16-1. F.P. also thanks the support of the MICINN Spanish grant AYA2010-21231-C02-01 and the Campus of International Excellence UAM+CSIC. A.K. acknowledges support from the NSF under a grant to NMSU.

REFERENCES

Abazajian, K. e. a. 2005, ApJ, 625, 613

- Angulo, R., Baugh, C. M., Frenk, C. S., Bower, R. G., Jenkins, A., & Morris, S. L. 2005, *MNRAS*, 362, L25
- Angulo, R. E., Baugh, C. M., Frenk, C. S., & Lacey, C. G. 2008, *MNRAS*, 383, 755
- Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, *ApJ*, 717, 379
- Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., Busha, M. T., Klypin, A. A., & Primack, J. R. 2011, *ArXiv e-prints*
- Berlind, A. A., & Weinberg, D. H. 2002, *ApJ*, 575, 587
- Blake, C., Collister, A., & Lahav, O. 2008, *MNRAS*, 385, 1257
- Blanton, M. R., Lin, H., Lupton, R. H., Maley, F. M., Young, N., Zehavi, I., & Loveday, J. 2003, *AJ*, 125, 2276
- Brown, M. J. I., Zheng, Z., White, M., Dey, A., Jannuzi, B. T., Benson, A. J., Brand, K., Brodwin, M., et al. 2008, *ApJ*, 682, 937
- Colless, M., et al. 2001, *MNRAS*, 328, 1039
- Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, *ApJ*, 647, 201
- Crocce, M., & Scoccimarro, R. 2006, *PRD*, 73, 063519
- . 2008, *PRD*, 77, 023533
- Cui, W., Liu, L., Yang, X., Wang, Y., Feng, L., & Springel, V. 2008, *ApJ*, 687, 738
- Davis, M., & Peebles, P. J. E. 1983, *ApJ*, 267, 465
- Eisenstein, D. J., Seo, H.-J., Sirko, E., & Spergel, D. N. 2007a, *ApJ*, 664, 675
- Eisenstein, D. J., Seo, H.-J., & White, M. 2007b, *ApJ*, 664, 660
- Eisenstein, D. J., et al. 2005, *ApJ*, 633, 560
- . 2011, *AJ*, 142, 72
- Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, *AJ*, 111, 1748
- Gottlöber, S., & Klypin, A. 2008, *ArXiv e-prints*
- Gunn, J. E., et al. 1998, *AJ*, 116, 3040
- . 2006, *AJ*, 131, 2332
- Guo, H., Zehavi, I., & Zheng, Z. 2011, *ArXiv e-prints*
- Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010, *MNRAS*, 404, 1111
- Hamilton, A. J. S. 1993, *ApJ*, 417, 19
- Hamilton, A. J. S., & Tegmark, M. 2004, *MNRAS*, 349, 115
- Ho, S., Cuesta, A., Seo, H.-J., de Putter, R., Ross, A. J., White, M., Padmanabhan, N., Saito, S., et al. 2012, *ArXiv e-prints*
- Jarosik, N., et al. 2011, *ApJS*, 192, 14
- Jing, Y. P. 2005, *ApJ*, 620, 559
- Kaiser, N. 1987, *MNRAS*, 227, 1
- Klypin, A., & Holtzman, J. 1997, *ArXiv Astrophysics e-prints*
- Klypin, A., Zhao, H., & Somerville, R. S. 2002, *ApJ*, 573, 597
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, 740, 102
- Knebe, A., et al. 2011, *MNRAS*, 415, 2293
- Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottlöber, S., Allgood, B., & Primack, J. R. 2004, *ApJ*, 609, 35
- Kravtsov, A. V., Klypin, A. A., & Khokhlov, A. M. 1997, *ApJS*, 111, 73
- Kulkarni, G. V., Nichol, R. C., Sheth, R. K., Seo, H.-J., Eisenstein, D. J., & Gray, A. 2007, *MNRAS*, 378, 1196
- Landy, S. D., & Szalay, A. S. 1993, *ApJ*, 412, 64
- Leauthaud, A., Tinker, J., Behroozi, P. S., Busha, M. T., & Wechsler, R. H. 2011, *ApJ*, 738, 45
- Li, C., & White, S. D. M. 2009, *MNRAS*, 398, 2177
- Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, *MNRAS*, 368, 715
- Manera, M., et al. 2012, in preparation
- Masjedi, M., Hogg, D. W., Cool, R. J., Eisenstein, D. J., Blanton, M. R., Zehavi, I., Berlind, A. A., Bell, E. F., et al. 2006, *ApJ*, 644, 54
- Masters, K. L., et al. 2011, *MNRAS*, 418, 1055
- Matsubara, T. 2008a, *PRD*, 78, 083519
- . 2008b, *PRD*, 77, 063530
- McBride, C., et al. 2009, in *Bulletin of the American Astronomical Society*, Vol. 41, American Astronomical Society Meeting Abstracts 213, 425
- Meiksin, A., White, M., & Peacock, J. A. 1999, *MNRAS*, 304, 851
- Montesano, F., Sánchez, A. G., & Phleps, S. 2010, *MNRAS*, 408, 2397
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
- Padmanabhan, N., White, M., Norberg, P., & Porciani, C. 2009, *MNRAS*, 397, 1862
- Phleps, S., Peacock, J. A., Meisenheimer, K., & Wolf, C. 2006, *A&A*, 457, 145
- Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E., & Primack, J. 2011, *ArXiv e-prints*
- Riebe, K., Partl, A. M., Enke, H., Forero-Romero, J., Gottlöber, S., Klypin, A., Lemson, G., Prada, F., et al. 2011, *ArXiv e-prints*
- Ross, A. J., & Brunner, R. J. 2009, *MNRAS*, 399, 878
- Ross, A. J., Percival, W. J., & Brunner, R. J. 2010, *MNRAS*, 407, 420
- Ross, A. J., et al. 2011, *MNRAS*, 417, 1350
- . 2012, in preparation
- Sánchez, A. G., Baugh, C. M., & Angulo, R. 2008, *MNRAS*, 390, 1470
- Schlaflly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, 737, 103
- Schlaflly, E. F., Finkbeiner, D. P., Schlegel, D. J., Jurić, M., Ivezić, Ž., Gibson, R. R., Knapp, G. R., & Weaver, B. A. 2010, *ApJ*, 725, 1175
- Scoccimarro, R., & Sheth, R. K. 2002, *MNRAS*, 329, 629
- Seo, H.-J., Eckel, J., Eisenstein, D. J., Mehta, K., Metchnik, M., Padmanabhan, N., Pinto, P., Takahashi, R., et al. 2010, *ApJ*, 720, 1650
- Skibba, R. A., & Sheth, R. K. 2009, *MNRAS*, 392, 1080
- Slosar, A., et al. 2011, *ArXiv e-prints*
- Swanson, M. E. C., Tegmark, M., Hamilton, A. J. S., & Hill, J. C. 2008, *MNRAS*, 387, 1391
- Taruya, A., Nishimichi, T., Saito, S., & Hiramatsu, T. 2009, *PRD*, 80, 123503
- Tasitsiomi, A., Kravtsov, A. V., Gottlöber, S., & Klypin, A. A. 2004, *ApJ*, 607, 125
- Tegmark et al., M. 2004, *ApJ*, 606, 702
- Trujillo-Gomez, S., Klypin, A., Primack, J., & Romanowsky, A. J. 2011, *ApJ*, 742, 16
- Vale, A., & Ostriker, J. P. 2004, *MNRAS*, 353, 189
- Wake, D. A., Croom, S. M., Sadler, E. M., & Johnston, H. M. 2008, *MNRAS*, 391, 1674
- Watson, D. F., Berlind, A. A., & Zentner, A. R. 2012, *ArXiv e-prints*
- Wetzel, A. R., & White, M. 2010, *MNRAS*, 403, 1072

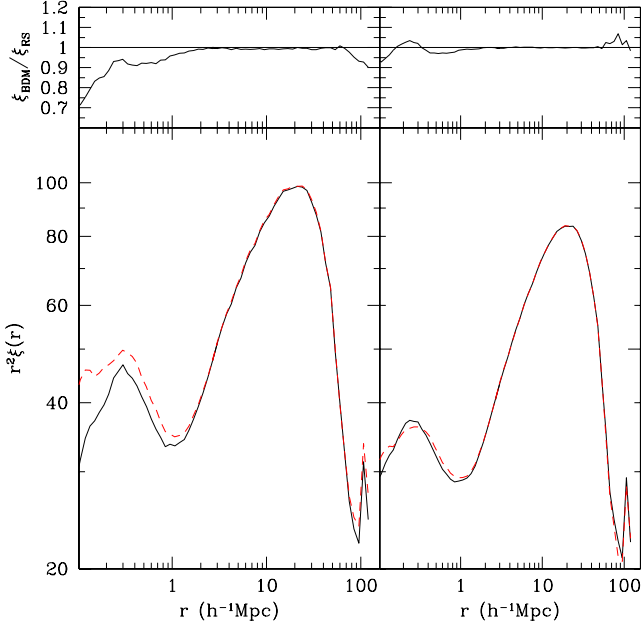


Figure A1. Comparison of the real-space correlation functions of (sub)haloes identified with the BDM and RockStar halo finders at $z = 0$ in the MultiDark simulation. Left panels are for (sub)haloes with maximum circular velocity $V_{\max} > 350 \text{ km s}^{-1}$ while right panels are for $V_{\max} > 300 \text{ km s}^{-1}$. Top panels present ratios of the correlation functions. Solid (dashed) lines in the bottom panels show the BDM (RockStar) correlation function multiplied by the square of radius.

- White, M., et al. 2011, *ApJ*, 728, 126
 York, D. G., et al. 2000, *AJ*, 120, 1579
 Zehavi, I., et al. 2002, *ApJ*, 571, 172
 —. 2005, *ApJ*, 630, 1
 —. 2011, *ApJ*, 736, 59
 Zentner, A. R., Berlind, A. A., Bullock, J. S., Kravtsov, A. V., & Wechsler, R. H. 2005, *ApJ*, 624, 505
 Zheng, Z., Berlind, A. A., Weinberg, D. H., Benson, A. J., Baugh, C. M., Cole, S., Davé, R., Frenk, C. S., et al. 2005, *ApJ*, 633, 791
 Zheng, Z., Zehavi, I., Eisenstein, D. J., Weinberg, D. H., & Jing, Y. P. 2009, *ApJ*, 707, 554

APPENDIX A: DEPENDENCE OF CLUSTERING ON DIFFERENT EFFECTS

The dependence of clustering with the halo finder used to identify virialized systems in the simulation is shown in Fig. A1 for the BDM (Klypin & Holtzman 1997; Riebe et al. 2011) and RockStar (Behroozi et al. 2011) codes. As an example we select all (sub)haloes present in the Multidark simulation at $z = 0$ with $V_{\max} > 300 \text{ km s}^{-1}$ and $V_{\max} > 350 \text{ km s}^{-1}$ in order to compute the real-space correlation function of the resulting halo catalogs. As can be seen in the figure the convergence between both halo finders is remarkable; for small ($\lesssim 1 h^{-1} \text{ Mpc}$) and large ($\gtrsim 70 h^{-1} \text{ Mpc}$) scales the difference in power is typically $\sim 10\%$.

To assess the clustering dependence with number den-

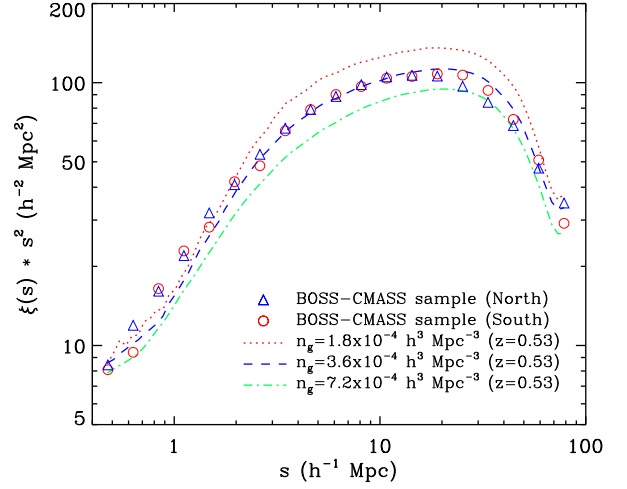


Figure A2. Redshift-space correlation function for different number densities of our MultiDark halo catalogs (including scatter) as indicated in the plot (see text). We compare these results with the DR9 BOSS-CMASS north and south galaxy sample in the redshift range $0.4 < z < 0.7$. For clarity the error bars are not shown.

sity we evaluate different correlations for the redshift-space and compare with observations (see Fig. A2). We compute three different correlation functions using the MultiDark halo catalogs at $z = 0.53$ assuming the stochasticity model presented in Section 3.2. In this way, we get the following number densities in each case: $n_g = [1.8, 3.6, 7.2] \times 10^{-4} h^3 \text{ Mpc}^{-3}$. The dashed line corresponds to our effective number density, i.e. $3.6 \times 10^{-4} h^3 \text{ Mpc}^{-3}$. As expected, doubling and dividing this value gives a weaker and stronger clustering signal respectively. In these extreme cases, the departure from observations is typically above observational uncertainties. This result reflects the importance of correctly matching the abundance of haloes to the one in observations. However, typical departures from the effective number density adopted in this work are smaller than 5% and do not appreciably change our final result.